Office de la propriété intellectuelle du Canada

Un organisme d'Industrie Canada

Canadian Intellectual Property Office

An Agency of Industry Canada

PCT / CA 03 01555

14 NOVEMBER 2003 14-11-03

*Bureau canadien des brevets*

*Certification*

*Canadian Patent Office*

*Certification*

La présente atteste que les documents ci-joints, dont la liste figure ci-dessous, sont des copies authentiques des documents déposés au Bureau des brevets.

This is to certify that the documents attached hereto and identified below are true copies of the documents on file in the Patent Office.

REC'D 21 JAN 2004

WIPO PCT

Specification and Drawings, as originally filed, with Application for Patent Serial No: **2,407,825**, on October 11, 2002, by **ANDREW J. SIMMONDS AND HENRY M. KRAUSE** for "Trap-Tagging: A Novel Method for the Identification and Purification of RNA-Protein Complexes".

**PRIORITY DOCUMENT**
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Agent certificateur/Certifying Officer

November 14, 2003

Date

Canada

(CIPO 68)
04-09-02

OPIC CIPO

# TRAP-Tagging: a novel method for the identification and purification of RNA-protein complexes.


## Abstract

With the recent completion of several genome sequencing projects, scientists have embarked on comprehensive attempts to unravel all of the interactions amongst their gene products. While many proteomics efforts are well under way, little attention has been paid to the RNA products of these genes. The variety and scope of roles that RNA molecules play in biological processes is only now beginning to be appreciated. Consequently, "ribonomics" will likely be the next focus of genomic efforts. Unfortunately, conventional methods for the isolation and identification of specific RNA-protein complexes are plagued by a number of problems not encountered in genomics or proteomics. Here we describe a method that circumvents these problems. The TRAP (Tandem RNA Affinity Purification) tag is a dual RNA tagging system that facilitates gentle purification of RNA molecules along with the proteins, RNAs and other small molecules specifically associated with them.

## Background

In addition to serving as essential intermediates between genes and proteins, RNA molecules also serve structural and regulatory roles in a rapidly growing list of biological processes. These include all of the basic steps of mRNA processing such as splicing, nuclear to cytoplasm transport, translation, and decay (Doudna and Rath, 2002; Erdmann et al., 2001; Pesole et al., 2001). Other known functions include the regulation of transcript initiation (Berkhout et al., 1989), dosage compensation (Bell et al., 1988; Lee and Jaenisch, 1997; Meller et al., 2000; Salido et al., 1992), telomere maintenance (Le et al., 2000) and DNA replication. Importantly, the genomes of many viruses are encoded as RNA rather than DNA, and much of their infective cycles are controlled by RNA biochemistry (Berkhout et al., 1989). Clearly, these molecules and processes are crucial for cell and pathogen viability, and are excellent targets for drug intervention.

A comprehensive dissection of the protein complexes that incorporate specific RNAs will help elucidate their functions, many of which are likely to be novel. However, the methodologies currently employed to identify RNA associated molecules are not ideally suited for such an endeavor. For example, RNA binding proteins generally do not have the same specificity as DNA binding proteins. Consequently, techniques that identify individual RNA-protein interactions frequently isolate proteins that are irrelevant to the processes being studied. Indeed, there is increasing evidence that many high affinity RNA/protein interactions require multiple contacts between several proteins and their cognate elements within the RNA molecule (Chartrand et al., 2001). An additional consequence of this complexity is that proteins being sought in extracts may already be bound in stable RNP complexes, allowing other abundant and relatively non-specific RNA binding proteins to bind RNA probes.

We have invented a method capable of isolating specific RNA-protein complexes formed in vivo.

**Summary**

This invention provides an isolated DNA construct comprising a transcription cassette, which comprises a promoter sequence, a bait sequence operably linked to the promoter, a transcriptional termination sequence which comprises a stop signal for RNA polymerase and a polyadenylation signal for polyadenylase, and at least two tag sequences. In one embodiment, the isolated DNA construct further comprises at least three insulator sequences. In another embodiment the isolated DNA construct comprises at least one streptavidin binding sequence [SEQ ID NO:1 SEQ ID NO:2] and at least one MS2 coat protein binding sequence [SEQ ID NO:4, SEQ ID NO:6 SEQ ID NO:7]. In yet another embodiment, the isolated DNA construct comprises at least one tag sequence which hybridizes to the streptavidin binding sequence [SEQ ID NO:2] and at least one tag sequence which hybridizes to the MS2 coat protein sequence [SEQ ID NO:4] under high stringency hybridization conditions.

The invention also provides an isolated DNA construct comprising a transcription cassette, which construct comprises, a promoter sequence, a bait sequence operably linked to the promoter, a transcriptional termination sequence, which comprises a stop signal for RNA polymerase and a polyadenylation signal for polyadenylase; and at least three tag sequences. In

one embodiment the isolated DNA construct further comprises at least four insulator sequences. In another embodiment the isolated DNA construct comprises at least one streptavidin binding sequence [SEQ ID NO:2] and at least two MS2 coat protein binding sequences [SEQ ID NO:7]. In yet another embodiment the isolated DNA construct at least one tag sequence which hybridizes to the streptavidin binding sequence [SEQ ID NO:2] and at least two tag sequences which hybridize to the MS2 coat protein sequence [SEQ ID NO:7] under high stringency hybridization conditions.

The present invention relates to a vector comprising an isolated DNA construct comprising a transcription cassette, which comprises a promoter sequence, a bait sequence operably linked to the promoter, a transcriptional termination sequence which comprises a stop signal for RNA polymerase and a polyadenylation signal for polyadenylase, and at least two tag sequences. In one embodiment, the isolated DNA construct further comprises at least three insulator sequences. In another embodiment the isolated DNA construct comprises at least one streptavidin binding sequence [SEQ ID NO:2] and at least one MS2 coat protein binding sequence [SEQ ID NO:4]. In yet another embodiment, the isolated DNA construct comprises at least one tag sequence which hybridizes to the streptavidin binding sequence [SEQ ID NO:2] and at least one tag sequence which hybridizes to the MS2 coat protein sequence [SEQ ID NO:4] under high stringency hybridization conditions.

The present invention also relates to a vector comprising an isolated DNA construct comprising a transcription cassette, which construct comprises, a promoter sequence, a bait sequence operably linked to the promoter, a transcriptional termination sequence, which comprises a stop signal for RNA polymerase and a polyadenylation signal for polyadenylase; and at least three tag sequences. In one embodiment the isolated DNA construct further comprises at least four insulator sequences. In another embodiment the isolated DNA construct comprises at least one streptavidin binding sequence [SEQ ID NO:2] and at least two MS2 coat protein binding sequences [SEQ ID NO:7]. In yet another embodiment the isolated DNA construct at least one tag sequence which hybridizes to the streptavidin binding sequence [SEQ ID NO:2] and at least two tag sequences which hybridize to the MS2 coat protein sequence [SEQ ID NO:7] under high stringency hybridization conditions.

The invention further provides a host cell transformed with a vector comprising an isolated DNA construct comprising a transcription cassette, which comprises a promoter sequence, a bait sequence operably linked to the promoter, a transcriptional termination sequence which comprises a stop signal for RNA polymerase and a polyadenylation signal for polyadenylase, and at least two tag sequences. In one embodiment, the isolated DNA construct further comprises at least three insulator sequences. In another embodiment the isolated DNA construct comprises at least one streptavidin binding sequence [SEQ ID NO:2] and at least one MS2 coat protein binding sequence [SEQ ID NO:4]. In yet another embodiment, the isolated DNA construct comprises at least one tag sequence which hybridizes to the streptavidin binding sequence [SEQ ID NO:1] and at least one tag sequence which hybridizes to the MS2 coat protein sequence [SEQ ID NO:4] under high stringency hybridization conditions.

The invention also provides for a host cell transformed with a vector The present invention also relates to a vector comprising an isolated DNA construct comprising a transcription cassette, which construct comprises, a promoter sequence, a bait sequence operably linked to the promoter, a transcriptional termination sequence, which comprises a stop signal for RNA polymerase and a polyadenylation signal for polyadenylase; and at least three tag sequences. In one embodiment the isolated DNA construct further comprises at least four insulator sequences. In another embodiment the isolated DNA construct comprises at least one streptavidin binding sequence [SEQ ID NO:2] and at least two MS2 coat protein binding sequences [SEQ ID NO:7]. In yet another embodiment the isolated DNA construct at least one tag sequence which hybridizes to the streptavidin binding sequence [SEQ ID NO:2] and at least two tag sequences which hybridize to the MS2 coat protein sequence [SEQ ID NO:7] under high stringency hybridization conditions.

Another aspect of the invention is an RNA fusion molecule comprising a target RNA sequence and at least two RNA tags, wherein at least one of the RNA tags interacts with a ligand in a reversible fashion. In one embodiment, the RNA fusion molecule further comprises at least three insulators. In another embodiment the RNA fusion molecule comprises at least one streptavidin binding tag [SEQ ID NO:3] and at least one MS2 coat protein binding tag [SEQ ID NO:5].

The current invention also relates to an RNA fusion molecule comprising a target RNA sequence and at least three RNA tags, wherein at least two of the RNA tags interact with a ligand in a reversible fashion. In one embodiment, the RNA fusion molecule further comprises at least 4 insulators. In another embodiment, the RNA fusion molecule comprises at least one streptavidin binding tag [SEQ ID NO:3] and at least two MS2 coat protein binding tags [SEQ ID NO:8].

The invention provides a method for isolating an RNA-protein complex formed *in vivo* comprising, expressing in a eukaryotic cell an RNA fusion molecule of the current invention, generating a whole cell extract, passing the extract over a first solid support comprising streptavidin protein, eluting a first eluate with the addition of biotin, collecting the first eluate, passing the first eluate over a second solid support comprising MS2 coat protein, eluting a second elute with the addition of a reagent selected from the group consisting of glutathione, RNAse or a denaturant, and collecting the second elute, wherein the second eluate contains the isolated RNA-protein complex.
The current invention provides a method of identifying a protein in an RNA-protein complex comprising isolating an RNA-protein complex formed *in vivo* comprising, expressing in a eukaryotic cell an RNA fusion molecule of the current invention, generating a whole cell extract, passing the extract over a first solid support comprising streptavidin protein, eluting a first eluate with the addition of biotin, collecting the first eluate, passing the first eluate over a second solid support comprising MS2 coat protein, eluting a second elute with the addition of a reagent selected from the group consisting of glutathione, RNAse or a denaturant, and collecting the second elute, wherein the second eluate contains the isolated RNA-protein complex and identifying the protein in the RNA-protein complex.

The invention also provides for a protein identified by isolating an RNA-protein complex formed *in vivo* comprising, expressing in a eukaryotic cell an RNA fusion molecule of the current invention, generating a whole cell extract, passing the extract over a first solid support comprising streptavidin protein, eluting a first eluate with the addition of biotin, collecting the first eluate, passing the first eluate over a second solid support comprising MS2 coat protein, eluting a second elute with the addition of a reagent selected from the group consisting of

glutathione, RNAse or a denaturant, and collecting the second elute, wherein the second eluate contains the isolated RNA-protein complex and identifying the protein in the RNA-protein complex.

Another aspect of the current invention is a method for isolating an RNA-protein complex formed *in vitro* comprising, (a) expressing a RNA fusion molecule of the current invention *in vitro*, (b) obtaining a whole cell extract, (c) passing the whole cell extract over a first solid support comprising streptavidin protein, (d) eluting a first eluate with the addition of biotin, (e) collecting the first eluate, (f) passing the first eluate over a second solid support comprising MS2 coat protein, (g) eluting a second elute with the addition of a reagent selected from the group consisting of glutathione, RNAse or a denaturant, and (h) collecting the second eluate, wherein the second eluate contains the isolated RNA-protein complex. In one embodiment steps (c) to (e) are repeated.

The current invention provides a method of identifying a protein in an RNA-protein complex comprising isolating an RNA-protein complex formed *in vitro* comprising (a) expressing a RNA fusion molecule of the current invention *in vitro*, (b) obtaining a whole cell extract, (c) passing the whole cell extract over a first solid support comprising streptavidin protein, (d) eluting a first eluate with the addition of biotin, (e) collecting the first eluate, (f) passing the first eluate over a second solid support comprising MS2 coat protein, (g) eluting a second elute with the addition of a reagent selected from the group consisting of glutathione, RNAse or a denaturant, and (h) collecting the second eluate, wherein the second eluate contains the isolated RNA-protein complex and identifying the protein in the RNA-protein complex. In one embodiment, steps (c) to (e) are repeated.

The invention also provides for a protein identified by isolating an RNA-protein complex formed *in vitro* comprising, (a) expressing a RNA fusion molecule of the current invention *in vitro*, (b) obtaining a whole cell extract, (c) passing the whole cell extract over a first solid support comprising streptavidin protein, (d) eluting a first eluate with the addition of biotin, (e) collecting the first eluate, (f) passing the first eluate over a second solid support comprising MS2 coat protein, (g) eluting a second elute with the addition of a reagent selected from the group

consisting of glutathione, RNAse or a denaturant, and (h) collecting the second eluate, wherein the second eluate contains the isolated RNA-protein complex and identifying the protein in the RNA-protein complex. In one embodiment, steps (c) to (e) are repeated.

The invention also relates to a method of screening for a compound that modulates the formation of an RNA-protein complex formed *in vivo* comprising, expressing in a eukaryotic cell an RNA fusion molecule of the current invention in the presence of a test compound, generating a whole cell extract, passing the extract over a first solid support comprising streptavidin protein, eluting a first eluate with the addition of biotin, collecting the first eluate, passing the first eluate over a second solid support comprising MS2 coat protein, eluting a second eluate with the addition or a reagent selected from the group consisting of glutathione, RNAse or a denaturant, collecting the second eluate, wherein the second eluate contains the isolated RNA-protein complex, measuring th amount of isolated RNA-protein complex present, and comparing the amount of isolated RNA-protein complex present in the absence of the compound to be tested.

The invention also provides for a method of screening for a compound that modulates the formation of an RNA-protein complex formed *in vitro* comprising, (a) expressing an RNA fusion molecule of the current invention *in vitro*, (b) obtaining a whole cell extract, (c) passing the whole cell extract over a first solid support comprising streptavidin protein, (d) eluting a first eluate with the addition of biotin, (e) collecting the first eluate, (f) passing the first eluate over a second solid support comprising MS2 coat protein, (g) eluting a second eluate with the addition of a reagent selected from the group consisting of glutathione, RNAse or a denaturant, (h) collecting the second eluate, wherein the second eluate contains the isolated RNA-protein complex, (i) measuring the amount of isolated RNA-protein complex present; and (j)comparing the amount of isolated RNA-protein complex present in the absence of the compound to be tested. In one embodiment, steps (c) to (e) are repeated.

The invention also provides for a kit for detecting RNA-protein complexes comprising an isolated DNA construct of the current invention.

The invention also provides for a kit for detecting RNA-protein complexes comprising a vector of the current invention.

**Brief Description of the Drawings**

Preferred embodiments of the invention will be described in relation to the drawings in which:

**Figure 1.** Tandem RNA affinity purification. A) RNAs of interest are tagged at their 5' or 3' end with two different RNA tags. The tagged RNAs are then expressed either in vitro or in vivo and tested for function. Functional complexes containing the tagged RNA are purified from extracts using two affinity resins, each of which is capable of binding one of the tags. An important aspect of the tags, particularly the first tag used, is that it must be capable of being dissociated from its affinity resin using conditions that do not disrupt the RNA-protein complex. Proteins eluted from the second resin are generally sufficiently pure for identification by SDS PAGE, silver staining, and Mass Spectrometry. Bound RNAs can also be identified using RTPCR or microarray analysis.

B) Sequence of the TRAP cassette. Sequences in parentheses indicate each of the different functional motifs within the TRAP cassette.

**Figure 2.** TRAP-tag purification using in vitro transcribed RNA.

A) In vitro purification of proteins from extracts. Embryonic whole cell extracts were mixed with TRAP-tagged constructs or control constructs \, and then passed over the two affinity columns. Eluates were subjected to SDS PAGE and silver staining. Lane 1: no RNA added to the extract. Lane 2: No bait RNA fused to the TRAP RNA. Lane 3: purification using TRAP RNA fused to a localization element from the 3'UTR of the *Drosophila wingless* gene mRNA (WLE1). . Lane 4: protein purification using TRAP RNA fused to a second transcript localizing element in the *wingless* mRNA 3' UTR (WLE2). Note that the RNAs containing the two baits (WLE1 and WLE2) bind proteins that are not bound by the resins or TRAP RNA alone. Interestingly, the proteins bound specifically by WLE1 and WLE2 also differ from each other.

B) In vitro purification of Bic-D from embryo extracts. Following the purification as described above, eluted proteins were subjected to SDS PAGE and then transferred to membranes for Western blotting with anti Bic-D antiserum. Lanes 1-4 are as described above. Note that the Bic-D signal is highly enriched in lanes 3 and 4 after TRAP purification with the WLE1 and WLE2

localization elements. Bic-D was only detectable in the crude extract after much longer exposures.

**Figure 3.** Localization of TRAP-tagged WLE RNAs in *Drosophila* embryos. To ensure that the TRAP-tag does not interfere with bait RNA function, WLE localization elements fused to TRAP RNAs were tested for localization activity in embryos. Panel A) shows the apical localization of fluorescently labeled WLE2 RNA after injection into a syncitial blastoderm stage embryo. Note the red RNA localized above the green labeled nuclei. Panel B) shows the random localization of a mutagenized WLE2 element that has no localizing activity. Panel C) shows apical localization of TRAP-tagged WLE2, showing localization that is indistinguishable from the untagged mRNA.

**Figure 4.** A)TRAP tag purification using RNA expressed in vivo. TRAP-tagged WLE1 and WLE2 localization elements were expressed in transgenic *Drosophila* embryos, and whole cell extracts were passed over tandem affinity columns. Lane 1: no TRAP-tagged RNA expressed. Lane 2: purification using TRAP-tagged WLE1. . Lane 3: purification using TRAP–tagged WLE2. Once again, proteins bound specifically by WLE1 and WLE2 differ.

**Table 1.** Suitability of tags for TRAP-tag purification. Tags used for affinity purification are shown in the left hand column. Sizes, affinity matrices, eluting reagents, and performance are shown in the columns to the right. Binding and elution efficiencies were determined using 32P-labeled RNAs expressed in vitro and are expressed as percentage of label loaded.

**Detailed Description**

The present invention will now be described more fully with reference to the accompanying drawings, in which preferred embodiments of the invention are shown. This invention may, however, be embodied in different forms and should not be construed as limited to the embodiments set forth herein. Rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the invention to those skilled in the art.

The term "bait sequence" as used herein, is a cDNA or DNA sequence that encodes a target RNA sequence. Examples of suitable bait sequences include RNAs, such as, the HIV Rev-

binding tat element, the *E. coli* N protein binding nut element, and various recognition elements within RNA splice sites.

The term “isolated DNA sequence” as described herein includes DNA whether single or double stranded. The sequence is isolated and/or purified (i.e. from its natural environment), in substantially pure or homogeneous form, free or substantially free of nucleic acid or genes of the species of interest or origin other than the promoter or promoter fragment sequence. The DNA sequence according to the present invention may be wholly or partially synthetic. The term "isolated" encompasses all these possibilities.

The term “operably linked” as described herein means joined as part of the same nucleic acid molecule, suitably positioned and oriented for transcription to be initiated from the promoter.

The term “promoter” as described herein refers to a sequence of nucleotides from which transcription may be initiated of DNA operably linked downstream (i.e. in the 3’ direction on the sense strand of double-stranded DNA). The promoter or promoter fragment may comprise one or more sequence motifs or elements conferring developmental and/or tissue-specific regulatory control of expression. For example, the promoter or promoter fragment may comprise a neural or gut-specific regulatory control element.

The term “DNA tag” as used herein refers to short DNA or cDNA sequences that encode a binding partner for a ligand. The ligand may be any molecule that specifically binds to the binding partner such as, antibiotics, antibodies or specific proteins. The DNA tags of the current invention may be located 3’ or 5’ to the bait sequence. DNA tags encode RNA tags.

The term “RNA tags” as used herein refers to short RNA sequences that function as a binding partner for a ligand. The RNA tags must be short, fully modular and must not interfere with each other or with the target RNA sequence. At least one of the RNA tags must interact with its binding partner in a reversible fashion.

The term "transcription cassette" as used herein refers to a nucleic acid sequence encoding a nucleic acid that is transcribed. To facilitate transcription, nucleic acid elements such as promoters, enhancers, transcriptional termination sequences and polyadenylation sequences are typically included in the transcription cassette.

The term "S1" as used herein refers to the streptavidin binding sequence as DNA [SEQ ID NO:1 or SEQ ID NO:2] or RNA [SEQ ID NO: 3]

The term "MS2" as used herein refers to MS2 coat protein binding sequence as DNA [SEQ ID NO: 4] or RNA [SEQ ID NO:5].

The term "2xMS2" as used herein refers to two MS2 coat protein binding sequences as DNA [SEQ ID NO:6 and SEQ ID NO:7] or RNA [SEQ ID NO:8]

The terminology used in the description of the invention herein is for the purpose of describing particular embodiments only, and is not intended to be limiting to the invention. As used in the description of the invention and the appended claims, the singular forms "a", "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. All publications, patent applications, patens and other references mentioned herein are incorporated by reference in their entirety.

The present invention relates to a method for isolating specific RNA-protein complexes formed *in vivo*. However, it can also be used to isolate or verify complexes formed in vitro.
In vivo complex formation and purification is accomplished by expressing tagged versions of the RNA of interest *in vivo* and then using the tag to isolate associated functional RNP complexes. Tags in the form of short RNA sequences that interact with specific proteins, antibiotics or synthetic ligands can be readily inserted 5' or 3' to the RNA of interest. Although a number of these potential RNA tags exist, purification with these tags gives at most a thousand-fold purification of the associated RNAs. By using two RNA tags, the TRAP-tag method of the

current invention provides approximately a million-fold purification of associated RNAs, which is sufficient for the identification of most cellular proteins. The tags in the current invention must be relatively short, fully modular, and must not interfere with each other or with the RNA of interest. In addition, at least one of the tags must interact with its ligand in a reversible fashion so that RNP complexes can be eluted intact from the first ligand matrix and bound to the second matrix (see Fig. 1A). When expressed in vivo, TRAP-tagged RNAs assemble into functional complexes, and these complexes are readily purified to homogeneity.

*Nucleic Acid Molecules*

*Functionally equivalent nucleic acid molecule or polypeptide sequence*

The term "isolated DNA sequence" refers to a DNA sequence the structure of which is not identical to that of any naturally occurring DNA sequence or to that of any fragment of a naturally occurring DNA sequence spanning more than three separate genes. The term therefore covers, for example, (a) DNA which has the sequence of part of a naturally occurring genomic DNA molecule; (b) a DNA sequence incorporated into a vector or into the genomic DNA of a prokaryote or eukaryote, respectively, in a manner such that the resulting molecule is not identical to any naturally occurring vector or genomic DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by reverse transcription of polyA RNA which can be amplified by PCR, or a restriction fragment; and (d) a recombinant DNA sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. Specifically excluded from this definition are nucleic acids present in mixtures of (i) DNA molecules, (ii) transfected cells, and (iii) cell clones, e.g., as these occur in a DNA library such as a cDNA or genomic DNA library.

Modifications in the DNA sequence, which result in production of a chemically equivalent or chemically similar amino acid sequence, are included within the scope of the invention. Modifications include substitution, insertion or deletion of nucleotides or altering the relative positions or order of nucleotides.

*Sequence identity*

The invention includes modified nucleic acid molecules with a sequence identity at least about: $\geq$95% to the DNA sequences provided in SEQ ID NO: 1 to SEQ ID NO: 4 (or a partial sequence

thereof or their complementary sequence). Preferably about 1, 2, 3, 4, 5, 6, to 10, 10 to 25, 26 to 50 or 51 to 100, or 101 to 250 nucleotides are modified. Sequence identity is most preferably assessed by the algorithm of the BLAST version 2.1 program advanced search (parameters as above). Blast is a series of programs that are available online at http//www.ncbi.nlm.nih.gov/BLAST.

References to BLAST searches are:

References to BLAST searches are:

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403_410.

Gish, W. & States, D.J. (1993) "Identification of protein coding regions by database similarity search." Nature Genet. 3:266_272.

Madden, T.L., Tatusov, R.L. & Zhang, J. (1996) "Applications of network BLAST server" Meth. Enzymol. 266:131_141.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI_BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389_3402.

Zhang, J. & Madden, T.L. (1997) "PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation." Genome Res. 7:649_656.

Other programs are also available to calculate sequence identity, such as Clustal W program (preferably using default parameters; Thompson, JD et al., Nucleic Acid Res. 22:4673-4680).

DNA sequences functionally equivalent to the S1 SEQ ID NO: 1, or MS2 SEQ ID NO: 3 can occur in a variety of forms as described above.

The sequences of the invention can be prepared according to numerous techniques. The invention is not limited to any particular preparation means. For example, the nucleic acid molecules of the invention can be produced by cDNA cloning, genomic cloning, cDNA synthesis, polymerase chain reaction (PCR) or a combination of these approaches (Current Protocols in Molecular Biology, F.M. Ausbel et al., 1989). Sequences may be synthesized using well-known methods and equipment, such as automated synthesizers.

*Hybridization*

Other functional equivalent forms of the S1 SEQ ID NO: 1 and SEQ ID NO: 2 and MS2 DNA SEQ ID NO: 3 and SEQ ID NO: 4 molecules can be isolated using conventional DNA-DNA or DNA-RNA hybridization techniques. These nucleic acid molecules and the S1 SEQ ID NO: 1 and SEQ ID NO: 2 and MS2 sequences can be modified without significantly affecting their activity.

The present invention also includes nucleic acid molecules that hybridize to one or more of the DNA sequences provided in SEQ ID NO:1 to SEQ ID NO:4 (or a partial sequence thereof or their complementary sequence). Such nucleic acid molecules preferably hybridize to all or a portion of S1 SEQ ID NO: 2 or MS2 SEQ ID NO: 4 or their complement under low, moderate (intermediate), or high stringency conditions as defined herein (see Sambrook et al. (most recent edition) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.; Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, NY)). The portion of the hybridizing nucleic acids is typically at least 15 (e.g. 20, 25, 30 or 50) nucleotides in length. The hybridizing portion of the hybridizing nucleic acid is at least 80% e.g. at least 95% or at least 98% identical to the sequence or a portion or all of a nucleic acid encoding S1 or S2 or their complement. Hybridizing nucleic acids of the type described herein can be used, for example, as a cloning probe, a primer (e.g. a PCR primer) or a diagnostic probe. Hybridization of the oligonucleotide probe to a nucleic acid sample typically is performed under stringent conditions. Nucleic acid duplex or hybrid stability is expressed as the melting temperature or Tm, which is the temperature at which a probe dissociates from a target DNA. This melting temperature is used to define the required stringency conditions. If sequences are to be identified that are related and substantially identical to the probe, rather than identical, then it is useful to first establish the lowest temperature at which only homologous hybridization occurs with a particular concentration of salt (e.g. SSC or SSPE). Then, assuming that 1% mismatching results in a 1 degree Celsius decrease in the Tm, the temperature of the final wash in the hybridization reaction is reduced accordingly (for example, if sequences having greater than 95% identity with the probe are sought, the final wash temperature is decreased by 5 degrees Celsius). In practice, the change in Tm can be between 0.5 degrees Celsius and 1.5 degrees Celsius per 1% mismatch. Low stringency conditions involve hybridizing at about: 1XSSC, 0.1% SDS at 50°C. High stringency conditions are: 0.1XSSC, 0.1% SDS at 65°C.

Moderate stringency is about 1X SSC 0.1% SDS at 60 degrees Celsius. The parameters of salt concentration and temperature can be varied to achieve the optimal level of identity between the probe and the target nucleic acid.

The present invention also includes nucleic acid molecules from any source, whether modified or not, that hybridize to genomic DNA, cDNA, or synthetic DNA molecules that encode. A nucleic acid molecule described above is considered to be functionally equivalent to a S1 nucleic acid molecule SEQ ID NO: 1 f the present invention if the sequence encoded by the nucleic acid molecule is recognized in a specific manner by streptavidin and is elutable by biotin. A nucleic acid molecule described above is considered to be functionally equivalent to a MS2 SEQ ID 4nucleic acid molecule of the present invention if the sequence encoded by the nucleic acid molecule is recognized in a specific manner by coat binding protein and is elutable by Glutathionine-S-Transferase (GST)-coat binding protein fusion protein.

*Vectors*

The present invention provides an expression vector comprising a transcription cassette. The transcription cassette can be cloned into a variety of vectors by means that are well known in the art. Such a vector may comprise a suitably positioned restriction site or other means for insertion of a transcription cassette. The vector may also contain a selectable marker. For use in an assay or experiment, commercially available vectors such as CMV Casper promoter vector may be employed. For use in gene therapy, vectors such as adenovirus, may be employed. Cell cultures transformed with the DNA sequences of the current invention are useful as research tools particularly for studies of RNA-protein complexes. One skilled in the art will appreciate that there are a wide variety of suitable vectors.

*Host Cells*

A further aspect of the present invention provides a host cell containing a transcription cassette of the current invention. Examples of particularly desirable host cells include ES, P19, COS, S2, SF9 cells. Methods known in the art for transformation, include but are not limited to electroporation, rubidium chloride, calcium chloride, calcium phosphate or chloroquine transfection, viral infection, phage transduction, microinjection, and the use of cationic lipid and lipid/amino acid complexes or of liposomes, or a large variety of other commercially available

and readily synthesized transfection adjuvants, are useful to transfer the vectors of the current invention into host cells. Host cells are cultured in conventional nutrient media. The media may be modified as appropriate for inducing promoters, amplifying nucleic acid sequences of interest or selecting transformants. The culture conditions, such as temperature, composition and pH will be apparent. After transformation, transformants may be identified on the basis of a selectable phenotype.

*RNA fusion molecules*

The current invention provides for RNA fusion molecules comprising RNA tags, insulator elements and target RNA sequences. A target RNA sequence may be an oligoribonucleotide sequence or a ribonucleic acid sequence. Generally, for use in this invention, the target RNA sequence is RNA, including ribosomal RNA, RNA encoded by a gene, messenger RNA, UTRs, ribozyme RNA, catalytic RNA, small nuclear RNA, small nucleolar RNA, etc., from a microorganism, or an RNA expressed by a cell infected with a virus, or RNA from a host cell, or RNA encoded by a genomic sequence; or RNA encoded by a chemically synthesized DNA sequence or random RNA encoded by randomly isolated DNA. Insulator elements may be placed on either side of the RNA tags and function to ensure proper folding of the RNA tags and to discourage interactions between the tags and the target RNA sequence. Examples of suitable insulator elements include, but are not limited to stretches of 4-5 identical nucleotides (eg 8-10 adenosines ) coupled with paired restriction sites that do not interact with the tag or bait sequences. The 5' and 3' restriction sites need to be identical as these sequences will hybridize forming a stem that forces the "insulator" polynucleoside sequences to be "unpaired" thus isolating the folded tag stem loop structure from the remainder of the RNA sequences produced from a specific vector. Insulator elements may also be called spacers.

*Method of Isolating*

The present invention relates to a method for isolating an RNA-protein complex formed *in vivo* comprising:

(a) expressing in a eukaryotic cell an RNA fusion molecule of the current invention,

(b) generating a whole cell extract;

(c) passing the extract over a first solid support comprising streptavidin protein;

(d) eluting a first eluate with the addition of biotin;

(e) collecting the first eluate;

(f) passing the first eluate over a second solid support comprising MS2 coat protein;

(g) eluting a second eluate with the addition of a reagent selected from the group consisting of glutathione, RNAse or a denaturant; and

(h) collecting the second eluate, wherein the second eluate contains the isolated RNA-protein complex.

The present invention also relates to a method for isolating an RNA-protein complex formed *in vitro* comprising:

(a) expressing the RNA fusion molecule of claim 11, 12, 13, 14, 15, or 16 *in vitro*;

(b) obtaining a whole cell extract;

(c) passing the whole cell extract over a first solid support comprising streptavidin protein;

(d) eluting a first eluate with the addition of biotin;

(a) collecting the first eluate;

(b) passing the first eluate over a second solid support comprising MS2 coat protein;

(c) eluting a second eluate with the addition of a reagent selected from the group consisting of glutathione, RNAse or a denaturant;and

(d) collecting the second eluate, wherein the second eluate contains the isolated RNA-protein complex.

The isolated protein part of the RNA-protein complex may then be identified by various methods and techniques including but not limited to SDS-page, silver staining, Western blotting and mass spectrometry.

Examples of suitable solid supports for use with the different embodiments of the current invention include affinity columns comprising bound streptavidin or bound MS2, wherein the MS2 can be bound to agarose or sepharose beads. MS2 affinity columns can also be made by crosslinking to resins such as affigel beads, or binding as a fusion protein to an appropriate resin (eg GST-MS2 to glutathione beads).

*Method of Screening*

The current invention relates to a method of screening for a compound that modulates the formation of an RNA-protein complex formed *in vivo* comprising:

(a) expressing in a eukaryotic cell an RNA fusion molecule of the instant invention, in the presence of a test compound;
(b) generating a whole cell extract;
(c) passing the extract over a first solid support comprising streptavidin protein;
(d) eluting a first eluate with the addition of biotin;
(e) collecting the first eluate;
(f) passing the first eluate over a second solid support comprising MS2 coat protein;
(g) eluting a second eluate with the addition of a reagent selected from the group consisting of glutathione, RNAse or a denaturant;
(h) collecting the second eluate, wherein the second eluate contains the isolated RNA-protein complex;
(i) measuring the amount of isolated RNA-protein complex present; and
(j) comparing the amount of isolated RNA-protein complex present in the absence of the compound to be tested.

Another embodiment of the current invention relates to a method of screening for a compound that modulates the formation of an RNA-protein complex formed *in vitro* comprising:

(a) expressing the RNA fusion molecule of claim 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17 *in vitro*;
(b) obtaining a whole cell extract;
(c) passing the whole cell extract over a first solid support comprising streptavidin protein;
(d) eluting a first eluate with the addition of biotin;
(e) collecting the first eluate;
(f) passing the first eluate over a second solid support comprising MS2 coat protein;
(g) eluting a second eluate with the addition of a reagent selected from the group consisting of glutathione, RNAse or a denaturant;
(h) collecting the second eluate, wherein the second eluate contains the isolated RNA-protein complex;
(i) measuring the amount of isolated RNA-protein complex present; and
(j) comparing the amount of isolated RNA-protein complex present in the absence of the compound to be tested.

Other assays (as well as variations of the above assays) will be apparent from the description of this invention. For example, the test compound may be either fixed or increased, a plurality of compounds or proteins may be tested at a single time. "Modulation" can refer to enhanced formation of the RNA-protein complex, a decrease in formation of the RNA-protein complex, a change in the type or kind of the RNA-protein complex or a complete inhibition of formation of the RNA-protein complex. Suitable compounds that may be used include but are not limited to proteins, nucleic acids, small molecules, hormones, antibodies, peptides, antigens, cytokines, growth factors, pharmacological agents including chemotherapeutics, carcinogenics, or other cells (i.e. cell-cell contacts). Screening assays can also be used to map binding sites on RNA or protein. For example, tag sequences encoding for RNA tags can be mutated (deletions, substitutions, additions) and then used in screening assays to determine the consequences of the mutations.

*Kits*

The invention includes kits for detecting RNA-protein complexes comprising at least one isolated DNA construct of the invention or at least one vector of the current invention.

*Tandem RNA purification*

A number of RNA motifs suitable as RNA affinity tags exist. We first tested five of these for potential use in our double-tagging system. These include the "streptotag", a streptomycin binding aptamer (Bachler et al., 1999), "S1",a streptavidin binding aptamer (Srisawat and Engelke, 2001), "D1", a sephadex binding aptamer (Srisawat et al., 2001), the MS2 phage coat protein binding RNA (Jurica et al., 2002) and the lambda phage nut RNA. Table 1 shows the relative binding and elution efficiencies of each $^{32}P$ –labeled tag and its ligand. Two of the five tags, the streptavidin (S1, SEQ ID NO: 1 and SEQ ID NO: 2) and MS2 coat protein (MS2) tags, were found to bind and elute efficiently under the desired purification conditions. Importantly, neither tag cross-reacted with any of the other tested ligands. Greater than 95% of the S1 tag SEQ ID NO: 1 and SEQ ID NO: 2 bound to streptavidin agarose beads, and could be recovered quantitatively with the addition of biotin. Approximately 80% of the loaded MS2 tag bound to GST-coat protein- beads, and approximately 70% of the loaded tag could be eluted with glutathione.

| RNA aptamer | SEQ ID NO | Length (nucleotides) | Affinity target | Eluted with: | % Bound | % Eluted |
|---|---|---|---|---|---|---|
| Streptotag | 9-DNA<br>10-RNA | 64 | 8-hydroxy-streptomycin | Streptomycin | 21% ± 2% | 12% ± 6% |
| MS2, 2xMS2 | 4,6-DNA<br>5,8-RNA | 38, 96 | Coat Binding Protein | Reduced Glutathione | 73% ± 3% | 68% ± 8% |
| S1 | 1- DNA<br>3- RNA | 68 | Streptavidin | Biotin | >99% | 94% ± 5% |
| D8 | 15 -DNA<br>16 - RNA | 64 | Sephadex | n/a | 34% ± 1% | 21% ± 10% |
| Nut 1-39 | 12- DNA<br>14 - RNA | 33 | N-protein 1-22 | n/a | <1% | <1% |

Table 1 – RNA aptamer tags tested for use in TRAP vectors

Next the ability of the Streptavidin and MS2 coat protein tags to function together and in the presence of an RNA target molecule was tested. Cassettes containing a T7 promoter, the two RNA tags, alternative target RNA insertion sites and a poly A tail were made (Figure 1B). Insulator elements, consisting of 8-10 Adenosines flanked by identical restriction sites, were placed on either side of each tag to ensure proper folding of the tags and to discourage interactions between the tags and the inserted target RNA. $^{32}$P-labeled RNAs were first tested for retention and elution on streptavidin and GST-coat protein columns. Both tags worked with much the same efficiency as when used individually. A construct containing 1 S1 tag SEQ ID NO: 1 and SEQ ID NO: 2 and 2 MS2 tags appeared to work best.

*TRAP tag purification using in vitro transcribed RNA*

Next the constructs were tested for the ability to purify specific RNA binding proteins from a complex protein mixture. Two, approximately 100-nucleotide long elements from the *Drosophila wingless* gene mRNA (WLE1 and WLE2) were chosen for this purpose. These elements are required for the asymmetrical localization of *wingless* transcripts to apical cytoplasm (Simmonds et al., 2001). The two elements show no similarity in sequence or predicted secondary structure and exhibit marked differences in their ability to localize transcripts. On the other hand, both appear to mediate localization via dynein-dependent

microtubule transport (Wilkie and Davis, 2001). Hence, they probably interact with unique but overlapping subsets of proteins.

The tagged RNAs were expressed in vitro, and the cold RNA mixed for 30 minutes with *Drosophila* embryo extracts prior to purification over the two columns. Figure 2A shows that each of the tagged localization elements did indeed associate with different subsets of proteins that were not bound by beads or tags alone. 9 of 19 proteins identified by Mass spectrometry are known or predicted RNA binding proteins (Simmonds and Krause, in preparation). Figure 2B shows that one of these proteins is Bic-D. Bic-D has been previously implicated as a protein required for apical mRNA transport in blastoderm stage *Drosophila* embryos (Bullock and Ish-Horowicz, 2001).

*Localization of TRAP-tagged WLE RNAs in Drosophila embryos*

The final test was to ensure that complexes formed on the tagged RNAs in vivo are both active and readily purified. To confirm this, tagged WLE constructs were first fluorescently labeled and injected into syncitial blastoderm stage embryos. RNAs with an apical localization motif will move from the site of injection upwards, between the syncitial nuclei to the apical surface (Bullock and Ish-Horowicz, 2001). Figure 3A shows untagged WLE2 RNA after localization to the apical surface. Figure 3C shows that TRAP-tagged WLE2 RNA localizes to the apical surface in an indistinguishable fashion. Thus, the tags appear to have no effect on the function of the localizing element. TRAP-tagged wingless localization elements expressed in transgenic embryos also localized apically (data not shown). Extracts were made from these transgenic fly lines and used for purification of WLE-associated proteins.

*TRAP tag purification using RNA expressed in vivo.*

Figure 4 shows that, as in vitro, each of the tagged WLE constructs binds a different subset of proteins. The identities of some of the se proteins were determined by Mass Spectrometry. Once again, one of the purified proteins included Bic-D.

Note that, although the proteins identified here were easily detected using a small amount of extract and silver staining, the reversibility of the two columns permits the optional use of a second round of purification to detect proteins of very low abundance and proteins that do not bind the bait stoichiometrically or in all cell types. The S1 tag SEQ ID NO:1 SEQ ID NO: 2 is

particularly well suited for a second round of purification. It provides high degrees of purification with little loss of material, and the biotin used for elution is easily removed. Biotin removal is achieved by running the eluate over an avidin column (the S1 tag SEQ ID NO:1 SEQ ID NO: 2 does not bind avidin). The flow-through is then bound to the second streptavidin column and eluted with biotin as before. This approach can also be used for prior removal of streptavidin binding proteins, should they be present in extracts in large amounts.

Clearly, this approach is applicable to any cell or tissue type. The TRAP cassette is simply placed into an appropriate vector. Although the in vivo application of the method is the most powerful version of this approach, in vitro assays are also clearly applicable. For example, using mutagenesis, the importance of specific nucleotides and structural aspects of known or newly discovered interactions can be rapidly tested with in vitro expressed RNAs and then confirmed in vivo. This approach is also amenable to high throughput analyses. This is particularly true for in vitro work with extracts, and with transfected or virally infected cells. With a little more effort, the approach can also be applied to transformed cells and transgenic tissues. For example, as has been done for proteins in yeast, TRAP tags could be placed within each yeast gene and substituted for the endogenous gene by homologous recombination. However, this approach is probably the most useful for small RNAs and functionally characterized RNA motifs. It is also possible to identify other RNAs bound within TRAP-purified complexes. This can be achieved either by RTPCR, or more globally by labeling the RNAsand hybridizing to microarrays.

Given the rapidly growing number of important processes controlled by RNAs and the proteins that bind them, TRAP-tagging should prove to be a key tool in the elucidation of these functions on a genomic scale. Once well characterized, functional RNA elements can serve as drug targets (RNAi etc). Viral RNAs such as HIV, hepatitis B, and the proteins that bind them, are particularly applicable targets.Examples of such uses include thetreatment of viral infections, the control of cellular proliferation and the stimulation of neuronal regeneration.

*Vector construction*

The initial TRAP vectors were constructed using a cassette-based approach to allow for maximum versatility and ease in transferring into specialized transgenesis vectors. The initial in-vitro expression vectors were constructed using the pSP72 cloning vector (Promega) as a backbone. This vector has a 5' T7 RNA polymerase site. The sequence for the S1 SEQ ID NO:1

SEQ ID NO: 2 and MS2 affinity tags were inserted using a pair of hybridized oligonucleotides. For the streptavidin aptamer the sequences were:

S1*Bgl*II5'- ATCTAAAAGACCGACCAGAATCATGCAAGTGCGTAAGATAGTCGCGGCCGGGAAAAAA

S13'- *Bgl*II3'-ATCTTTTTCCCGGCCCGCGACTATCTTACGCACTTGCATGATTCTGGTCGGTCTTTTA

that produce a S1 cassette: SEQ ID NO: 1 when inserted into the BglII site of pSP72 (See Figure 2). The MS2 aptamer was created from four linked oligonucleotides MS2#1 5' CAAACGACTCTAGAAAACATGAGGATCACCCATGTCTGCAGG

MS2#1 3' TCGACCTGCAGACATGGGTGATCCTCATGTTTTCTAGAGTCGTTTTTGAGC

MS2 #2 5' TCGACTCTAGAAACATGAGGATCACCCATGTCTGCAGGTCAAAAAGAGCT and MS2 #2 3' CTTTTTGACCTGCAGACATGGGTGATCCTCATGTTTTCTAGAG that form a cassette containing two MS2 hairpins SEQ ID NO:6 when cloned into the pSP72 *Sac*II site. This vector was then sequenced to ensure that the aptamer sequences were in the correct orientation. Other primers used to create the other tags tested included: 5'Streptotag KpnI

CAAAAGGATCGCATTTGGACTTCTGCCCAGGGTGGCACCACGTGCGGATCCAAAAGGTAC

3' Streptotag KpnI

CTTTTGGATCCGACCGTGGTGCCACCCTGGGCAGAAGTCCAAATGCGATCCTTTTGGTAC which when hybridized and cloned into the KpnI site of pSP72 produce the Streptotag cassette SEQ ID NO: 9,

N-5' KpnI – GATCCTTTTCGGGTGAAAAAGGGCTTTTG N3' KpnI

GATCCAAAAGCCCTTTTTCAGGGCAAAG. that when hybridized and cloned into the KpnI site of pSP72 produce the Nut cassette SEQ ID NO: 12. Also a pre-made cassette D8 encoding a Sephadex binding hairpin was also tested SEQ ID NO: 15 (Srisawat et al., 2001).

The subsequent vectors pTRAPS1, pTRAPMS2, pTRAPS1MS2, pTRAPN, pTRAPS1N pTRAPS1D8, and pTRAP D8 all contain several sites for cloning bait sequences, (pTRAPS1MS2 is shown in Figure 2).

To create in-vitro labeled RNA, the pTRAP vector was linearized by cutting to completion with XhoI. This cut DNA was treated with phenol and chloroform to remove the restriction enzyme. A 25µl RNA transcription reaction contains: 1µg of linearized pTRAP DNA, 5µl of 5x T7 RNA polymerase buffer (400mM Tris-HCl pH 8.0, 60mM $MgCl_2$), 5µl 10mM NTP mix, 1µl 0.75mM Dithiothreitol (RNAse free), 20U of placental RNAse inhibitor (MBI), 15U of T7 RNA polymerase and RNAse free water to 25µl. This reaction was incubated at 37°C for 2 hours and then the

*Purification of GST-CP beads*

A coat protein GST fusion protein was made by subcloning a PCR fragment consisting of the entire open reading frame of coat protein gene, (with a BamHI restriction site added 3' and XhoI added 5') into the pGEX4T vector (Pharmacia). The GST fusion protein was expressed in *E. coli* BL21 cells grown at 37°C for 3 hours ($OD_{600}$ of 1.8) and then induced with100mM IPTG. for 4.5 hours. Cells were pelleted in 250 ml aliquots, quick frozen in liquid nitrogen and stored for as long as 2 months at –70°C. Cell pellets were lysed by sonication (5 min at 50%) and bound to Glutathione–Sepharose beads (Pharmacia) following the manufactures directions. Purified beads can be stored in PBS for up to 1 month at 4 °C or alternatively the purified GST-Coat Protein can be eluted using reduced Glutathione (Sigma). The purified protein is then concentrated in a centricon filter (Millipore) and finally re-constituted in 1x PBS with 10% glycerol. This solution can be stored at –70°C for more than six months. For use in TRAP purification, the purified protein can be re-bound to GST-sepharose (5 mg/ml) or alternatively, can be pre-bound to the RNA and then bound to the affinity matrix.

*Transgenic Lines*

To create transgenic Drosophila, the *BglII-PvuII* fragement of pTRAPS1MS2+WLE1, pTRAPS1MS2+WLE1 or pTRAPS1MS2 (no insert) was cloned into a *BglII-StuI* site within the transposable element based pCASPER-HS vector (Thummel and Pirrotta, 1992). These vectors pCASPER-TRAP-WLE1, pCASPER-TRAP-WLE2 and pCASPER-TRAP were then introduced into Drosophila embryos by microinjection (Spradling and Rubin, 1982). For each construct, at least three independent transgenic lines were isolated.

*Extract preparation*

TRAP Purification Buffer (TPB) was used for all steps of the purification including isolation of the extract (5x stock solution = 300mM HEPES pH 7.4, 50mM MgCl, 400mM NaCl, 0.5% Triton X-100). TPB working solution is made by diluting the 5x stock and adding proteinase inhibitor (Complete, EDTA free; Roche) and 0.3mM DTT. *Drosophila* embryos were collected for 4 or 12 hours and aged an additional 4 hours. For each transgene, the level of transgenic RNA was determined empirically by semi-quantitative RT-PCR to allow for RNA expression that was similar in level to the endogenous transcript. For example, TRAP constructs containing the WLE

bait sequences, the transgenic embryos were induced using a 30 min heat pulse (36.5°C) and allowed to recover for 20 minutes. Following dechorionation, transfer embryos to a chilled dounce homogenizer. All further steps are at 4°C. Add enough TBP to cover embryos and homogenize using 10 strokes with a loose (A) and 10 strokes with a tight (B) pestle. Transfer homogenate to RNAse free 1.5ml tubes and spin 10 minutes at 14000g. Transfer supernatant to a new tube and repeat until extract is clear (avoid lipid layer above the extract). Add an additional 10% glycerol and freeze in liquid nitrogen. Lysates prepared in this way are approximately 5mg/ml protein and can be stored for up to 3 months at –70°C.

*TRAP purification*

RNAse free conditions and solutions made with DEPC treated water were used throughout. Biotin-related proteins were first removed from the extract by mixing the extract with Avidin agarose beads (Sigma). For each 500μl of extract, 100μl settled volume of Avidin beads is washed 3 times in 800 μl TBP and then incubated with thawed lysate for 10 minutes at 4°C. The avidin beads are then removed by passing the mixture over an RNAse free mini chromatography column (Bio Rad). Eluates are collected and mixed with 50 μl/ml extract, pre-equilibrated streptavidin agarose beads (Sigma). After gentle rocking for 1h at 4°C, the mixture is added to a plugged mini column and allowed to settle. After elution, columns are washed three times with 800μl TBP. The bound RNA/protein complexes are then eluted by the addition of Biotin. 250μl of Biotin elution buffer, (1x TBP + 5mM Biotin, Sigma). Incubate for 1hr at 4°C. Collect the eluate, wash the column once with an additional 250μl Biotin elution buffer and pool the wash with the first eluate. An option at this point is to repurify the eluate over a second streptavidin column by removing the biotin (using Avidin-agarose beads as described above) and repeating the procedure.

The streptavidin eluate is then bound to the GST-CP beads (described above). Equilibrate 100μl of a 50% slurry of GST-CP sepharose beads per 500μl of streptavidin eluate in 1xTBP. Add the streptavidin column eluate and rock for 1h at 4°C. Pour into a plugged mini column, let beads settle and then let lysate flow through. Wash three times with 500μl of 1xTBP. As above, a Bio-Rad or Pierce protein assay can be used to determine total number of washes needed.

The bound RNA and proteins can be eluted using glutathione, high salt, RNAse or various denaturants (*eg*. urea, SDS). The column is capped and one bed volume of elution buffer added.

For Glutathione and RNAse elutions, the mixture is rocked for 1 hr at 4°C. The Glutathione elution buffer is as described by Pharmacia, and RNAse elution buffer includes a 200µl of 2mg/mlRNAseA and 5000u/mlRNAse T1 (Fermentas) in 1ml of RNAse buffer (10mM Tris-HCl pH 7.5 and 10mM $MgCl_2$). Wash the column three times with an equal amount of appropriate buffer and pool the eluates. Proteins are then resolved by SDS PAGE and identified by Trypsin proteolysis, Mass-spectromety (Fenyo et al., 1998) and submission of the data to genomic databases.

References

Bachler, M., Schroeder, R., and von Ahsen, U. (1999). StreptoTag: a novel method for the isolation of RNA-binding proteins. RNA *5*, 1509-1516.

Bell, L. R., Maine, E. M., Schedl, P., and Cline, T. W. (1988). Sex-lethal, a Drosophila sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. Cell *55*, 1037-1046.

Berkhout, B., Silverman, R. H., and Jeang, K. T. (1989). Tat trans-activates the human immunodeficiency virus through a nascent RNA target. Cell *59*, 273-282.

Bullock, S. L., and Ish-Horowicz, D. (2001). Conserved signals and machinery for RNA transport in Drosophila oogenesis and embryogenesis. Nature *414*, 611-616.

Chartrand, P., Singer, R. H., and Long, R. M. (2001). RNP localization and transport in yeast. Annu Rev Cell Dev Biol *17*, 297-310.

Doudna, J. A., and Rath, V. L. (2002). Structure and function of the eukaryotic ribosome: the next frontier. Cell *109*, 153-156.

Erdmann, V. A., Barciszewska, M. Z., Hochberg, A., de Groot, N., and Barciszewski, J. (2001). Regulatory RNAs. Cell Mol Life Sci *58*, 960-977.

Fenyo, D., Qin, J., and Chait, B. T. (1998). Protein identification using mass spectrometric information. Electrophoresis *19*, 998-1005.

Le, S., Sternglanz, R., and Greider, C. W. (2000). Identification of two RNA-binding proteins associated with human telomerase RNA. Mol Biol Cell *11*, 999-1010.

Lee, J. T., and Jaenisch, R. (1997). Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature *386*, 275-279.

Meller, V. H., Gordadze, P. R., Park, Y., Chu, X., Stuckenholz, C., Kelley, R. L., and Kuroda, M. I. (2000). Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in *Drosophila*. Curr Biol *10*, 136-143.

Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., and Liuni, S. (2001). Structural and functional features of eukaryotic mRNA untranslated regions. Gene *276*, 73-81.

Salido, E. C., Yen, P. H., Mohandas, T. K., and Shapiro, L. J. (1992). Expression of the X-inactivation-associated gene XIST during spermatogenesis. Nat Genet *2*, 196-199.

Simmonds, A. J., dosSantos, G., Livne-Bar, I., and Krause, H. M. (2001). Apical localization of *wingless* transcripts is required for Wingless signaling. Cell *105*, 197-207.

Spradling, A. C., and Rubin, G. M. (1982). Transposition of cloned P elements into *Drosophila* germ line chromosomes. Science *218*, 341-347.

Srisawat, C., and Engelke, D. R. (2001). Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. RNA *7*, 632-641.

Srisawat, C., Goldstein, I. J., and Engelke, D. R. (2001). Sephadex-binding RNA ligands: rapid affinity purification of RNA from complex RNA mixtures. Nucleic Acids Res *29*, E4.

Thummel, C. S., and Pirrotta, V. (1992). Technical Notes: pCaSpeR P-element vectors. Drosophila Information Service *71*.

Wilkie, G. S., and Davis, I. (2001). *Drosophila wingless* and pair-rule transcripts localize apically by dynein-mediated transport of RNA particles. Cell *105*, 209-219.

SEQUENCE LISTING

<110> Krause, Henry

Simmonds, Andrew

<120> TRAP-Tagging: a novel method for the identification and purification of RNA-protein complexes

<130> 3110 0023

<160> 16

<170> PatentIn version 3.0

<210> 1

<211> 68

<212> DNA

<213> synthetic construct

<400> 1
atcgataaaa agaccgacca gaatcatgca agtgcgtaag atagtcgcgg gccgggaaaa 60

aaatcgat 68

<210> 2

<211> 45

<212> DNA

<213> synthetic construct

<400> 2
gaccgaccag aatcatgcaa gtgcgtaaga tagtcgcggg ccggg 45

<210> 3

<211> 68

<212> RNA

<213> synthetic construct

<400> 3
aucgauaaaa agaccgacca gaaucaugca agugcguaag auagucgcgg gccgggaaaa 60

aaaucgau 68

<210> 4

<211> 38

<212> DNA

<213> synthetic construct

<400> 4
gactctagaa acatgaggat cacccatgtc tgcaggtc 38

<210> 5

<211> 38

<212> RNA

<213> synthetic construct

<400> 5
gacucuagaa acaugaggau cacccauguc ugcagguc 38

<210> 6

<211> 96

<212> DNA

<213> synthetic construct

<400> 6
gagctcaaaa acgactctag aaacatgagg atcacccatg tctgcaggtc gactctagaa 60

acatgaggat accatgtctg caggtcaaaa gagctc 96

<210> 7

<211> 75

<212> DNA

<213> synthetic construct

<400> 7
cgactctaga aacatgagga tcacccatgt ctgcaggtcg actctagaaa catgaggata 60

ccatgtctgc aggtc 75

<210> 8

<211> 96

<212> RNA

<213> synthetic construct

<400> 8
gagcucaaaa acgacucuag aaacaugagg aucacccaug ucugcagguc gacucuagaa 60

acaugaggau accaugucug caggucaaaa gagcuc 96

<210> 9

<211> 64

<212> DNA

<213> synthetic construct

<400> 9
gtaccaaaag gatcgcattt ggacttctgc ccagggtggc accacgtgcg gatccaaaag 60

gtac 64

<210> 10

<211> 46

<212> DNA

<213> synthetic construct

<400> 10
ggatcgcatt tggacttctg cccagggtgg caccacgtgc ggatcc 46

<210> 11

<211> 64

<212> RNA

<213> synthetic construct

<400> 11
guaccaaaag gaucgcauuu ggacuucugc ccaggguggc accacgugcg gauccaaaag 60

guac 64

<210> 12

<211> 33

<212> DNA

<213> synthetic construct

<400> 12
gatccttttc gggtgaaaaa gggcttttgg tac 33

<210> 13

<211> 15

<212> DNA

<213> synthetic construct

<400> 13
cgggtgaaaa agggc 15

<210> 14

<211> 33

<212> RNA

<213> synthetic construct

<400> 14
gauccuuuuc gggugaaaaa gggcuuuugg uac 33

<210> 15

<211> 64

<212> DNA

<213> synthetic construct

<400> 15
ccgaccagaa gtccgagtaa tttacgtttt gatacggttg cggaacttgc tatgtgcgtc 60

taca 64

<210> 16

<211> 64

<212> RNA

<213> synthetic construct

<400> 16
ccgaccagaa guccgaguaa uuuacguuuu gauacgguug cggaacuugc uaugugcguc 60

uaca 64

We claim:

1. An isolated DNA construct comprising a transcription cassette, which construct comprises:
(a) a promoter sequence;
(b) a bait sequence operably linked to the promoter;
(c) a transcriptional termination sequence, which comprises a stop signal for RNA polymerase and a polyadenylation signal for polyadenylase; and
(d) at least two tag sequences.

2. The isolated DNA construct of claim 1 further comprising at least three insulator sequences.

3. The isolated DNA construct of claim 1 or 2 wherein the tag sequences comprise at least one streptavidin binding sequence [SEQ ID NO:2] and at least one MS2 coat protein binding sequence [SEQ ID NO:4].

4. The isolated DNA construct of claim 1 or 2 wherein the tag sequences comprise at least one sequence which hybridizes to streptavidin binding sequence [SEQ ID NO:2] and at least one sequence which hybridizes to MS2 coat protein sequence [SEQ ID NO:4] under high stringency hybridization conditions.

5. An isolated DNA construct comprising a transcription cassette, which construct comprises:
(a) a promoter sequence;
(b) a bait sequence operably linked to the promoter;
(c) a transcriptional termination sequence, which comprises a stop signal for RNA polymerase and a polyadenylation signal for polyadenylase; and
(d) at least three tag sequences.

6. The isolated DNA construct of claim 5 further comprising at least four insulator sequences.

7. The isolated DNA construct of claim 5 or 6 wherein the tag sequences comprise at least one streptavidin binding sequence [SEQ ID NO:2] and at least two MS2 coat protein binding sequences [SEQ ID NO:7].

8. The isolated DNA construct of claim 5 or 6 wherein the tag sequences comprise at least one sequence which hybridizes to streptavidin binding sequence [SEQ ID NO:2] and at least two sequences which hybridizes to MS2 coat protein sequence [SEQ ID NO:7] under high stringency hybridization conditions.

9. A vector comprising the isolated DNA construct of claim 1, 2, 3, 4, 5, 6, 7, or 8.

10. A host cell transformed with the vector of claim 9.

11. An RNA fusion molecule comprising:
(a) a target RNA sequence; and
(b) at least two RNA tags, wherein at least one of the RNA tags interacts with a ligand in a reversible fashion.

12. The RNA fusion molecule of claim 11 further comprising at least three insulators.

13. The RNA fusion molecule of claim 11 or 12 wherein the RNA tags comprise at least one streptavidin binding tag [SEQ ID NO:3] and at least one MS2 coat protein binding tag [SEQ ID NO:5].

14. An RNA fusion molecule comprising;
(a) a target RNA sequence; and
(b) at least three RNA tags, wherein at least two of the RNA tags interacts with a ligand in a reversible fashion.

15. The RNA fusion molecule of claim 14 further comprising at least 4 insulators.

16. The RNA fusion molecule of claim 14 or 15 wherein the RNA tags comprise at least one streptavidin binding tag [SEQ ID NO:2] and at least two MS2 coat protein binding tags [SEQ ID NO:7].

17. A method for isolating an RNA-protein complex formed *in vivo* comprising:
(a) expressing in a eukaryotic cell the RNA fusion molecule of claim 11, 12, 13, 14, 15, or 16;
(b) generating a whole cell extract;
(c) passing the extract over a first solid support comprising streptavidin protein;
(d) eluting a first eluate with the addition of biotin;
(e) collecting the first eluate;
(f) passing the first eluate over a second solid support comprising MS2 coat protein;
(g) eluting a second eluate with the addition of a reagent selected from the group consisting of glutathione, RNAse or a denaturant; and
(h) collecting the second eluate, wherein the second eluate contains the isolated RNA-protein complex.

18. A method of identifying a protein in an RNA-protein complex comprising the method of isolating an RNA-protein complex according to the method of claim 17 and identifying the protein in the RNA-protein complex.

19. A protein identified by the method of claim 18.

20. A method for isolating an RNA-protein complex formed *in vitro* comprising:
(a) expressing the RNA fusion molecule of claim 11, 12, 13, 14, 15, or 16 *in vitro*;
(b) obtaining a whole cell extract;
(c) passing the whole cell extract over a first solid support comprising streptavidin protein;
(d) eluting a first eluate with the addition of biotin;
(e) collecting the first eluate;

(f) passing the first eluate over a second solid support comprising MS2 coat protein;

(g) eluting a second eluate with the addition of a reagent selected from the group consisting of glutathione, RNAse or a denaturant;and

(h) collecting the second eluate, wherein the second eluate contains the isolated RNA-protein complex.

21. The method of claim 20, wherein the steps (c) to (e) are repeated.

22. A method of identifying a protein in an RNA-protein complex comprising the method of isolating an RNA-protein complex according to the method of claim 20 and identifying the protein in the RNA-protein complex.

23. A protein identified by the method claim 22.

24. A method of screening for a compound that modulates the formation of an RNA-protein complex formed *in vivo* comprising:

(a) expressing in a eukaryotic cell the RNA fusion molecule of claim 11, 12, 13, 14, 15 or 16, in the presence of a test compound;

(b) generating a whole cell extract;

(c) passing the extract over a first solid support comprising streptavidin protein;

(d) eluting a first eluate with the addition of biotin;

(e) collecting the first eluate;

(f) passing the first eluate over a second solid support comprising MS2 coat protein;

(g) eluting a second eluate with the addition of a reagent selected from the group consisting of glutathione, RNAse or a denaturant;

(h) collecting the second eluate, wherein the second eluate contains the isolated RNA-protein complex;

(i) measuring the amount of isolated RNA-protein complex present; and

(j) comparing the amount of isolated RNA-protein complex present in the absence of the compound to be tested.

25. A method of screening for a compound that modulates the formation of an RNA-protein complex formed *in vitro* comprising:

(a) expressing the RNA fusion molecule of claim 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17 *in vitro*;

(b) obtaining a whole cell extract;

(c) passing the whole cell extract over a first solid support comprising streptavidin protein;

(d) eluting a first eluate with the addition of biotin;

(e) collecting the first eluate;

(f) passing the first eluate over a second solid support comprising MS2 coat protein;

(g) eluting a second eluate with the addition of a reagent selected from the group consisting of glutathione, RNAse or a denaturant;

(h) collecting the second eluate, wherein the second eluate contains the isolated RNA-protein complex;

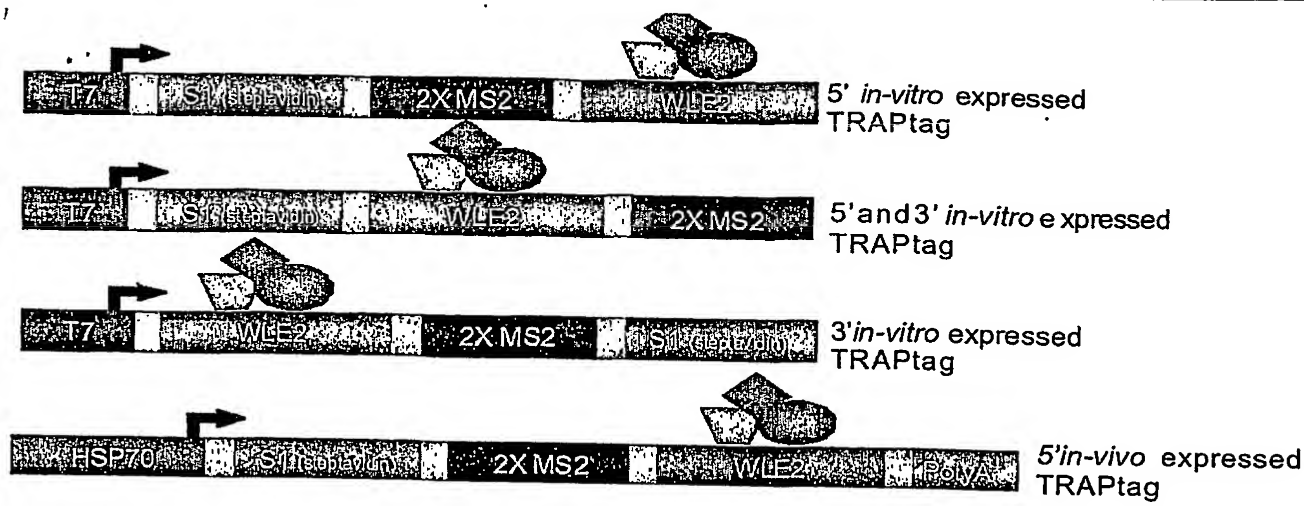
(i) measuring the amount of isolated RNA-protein complex present; and

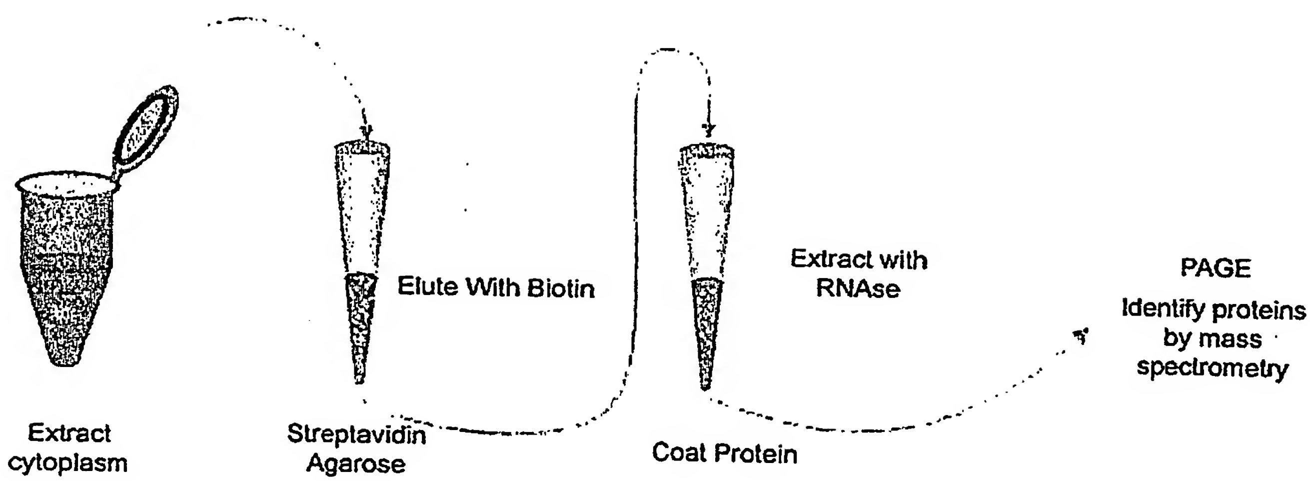
(j) comparing the amount of isolated RNA-protein complex present in the absence of the compound to be tested.

26. The method of claim 25, wherein steps (c) to (e) are repeated.

27. A kit for detecting RNA-protein complexes comprising the isolated DNA construct of claim 1, 2, 3, 4, 5, 6, 7, or 8.

28. A kit for detecting RNA-protein complexes comprising the vector of claim 9.




S1 (Streptavidin Binding) Aptamer Tag

GATC TGATGTCATC GAT AAAAAA GACCGACCAGAATCATGCAAGTGCGTAAGATAGTCGCGGGCCGGGA AAAAA ATCGAT

*BglII* *ClaI* *spacer* *spacer* *ClaI*

2xMS2 (Coat Protein Binding) Aptamer Tag

AATT CGAG CTC AAAAAA CGACTCTAGAAACATGAGGATCACCCATGTCTGCAGGTCGACTCTAGAAACATGAGGATACCA

*EcoRI* *SstI* *spacer*

TCTGCAGGTC AAAA GAGCTC GGTACCCGGGGATCC

*spacer* *SstI* *KpnI* *BamHI*